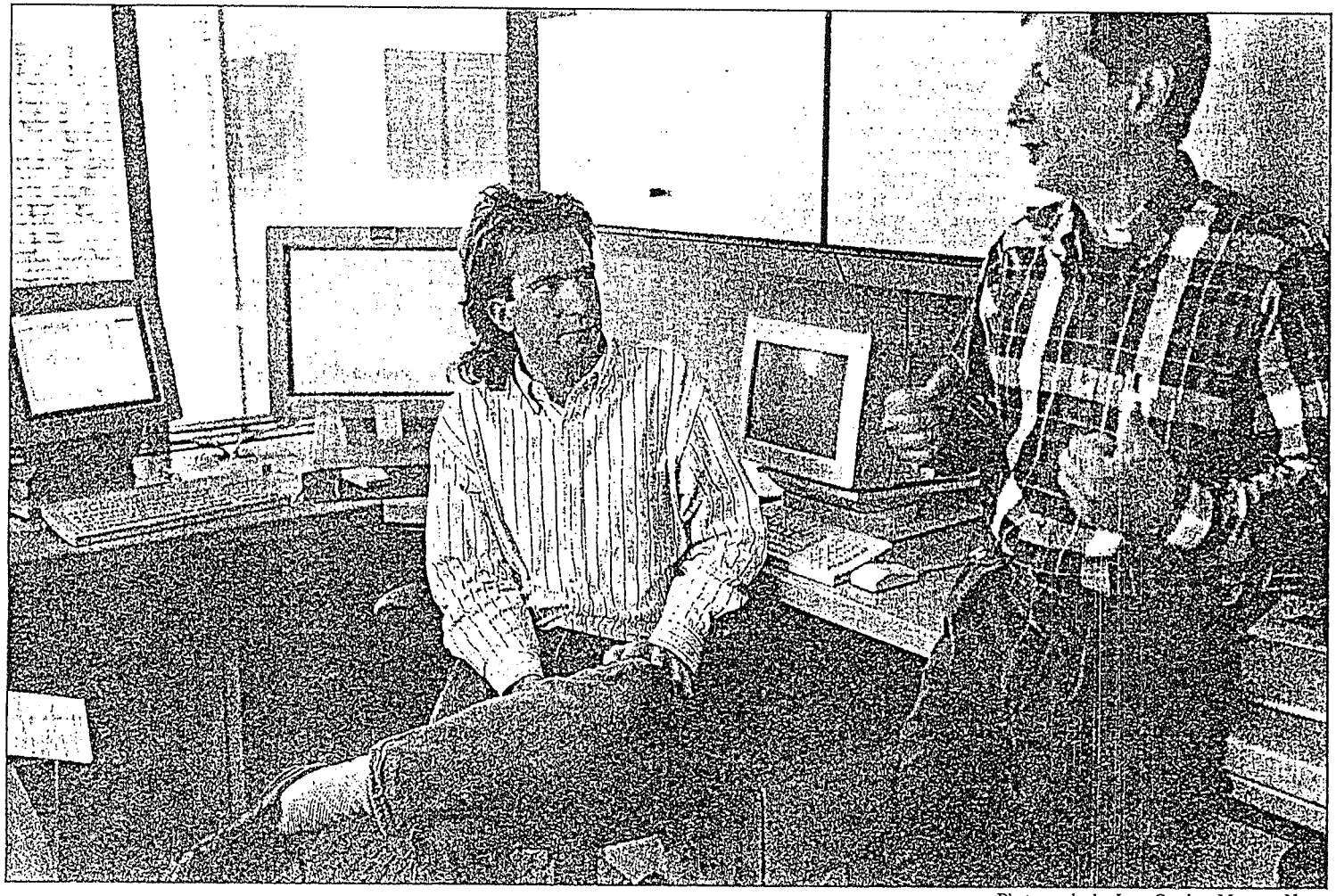
San Jose Mercury News, Sunday, July 21, 1991

The world at your keyboard



Photographs by Lara Cerri — Mercury News

Kevin Tiene listens as Harvey Lehtman explains Apple's Rosebud on-line library project, below

Network to unite data bases

By John Markoff
New York Times

The development of a nationwide data network will allow personal computer users to tap sources as large as the Library of Congress or receive their own personalized electronic newspapers.

Several innovations, taken together, have demonstrated that searching vast computer data bases can be easier than consulting a card catalog and not nearly as difficult or expensive as computer searches are today.

Computer users might read some Dickens more readily than they could check out David Copperfield from the local library.

Those in the industry say that users with little computer skills will soon be able to search through several terabytes of information, or several trillion characters of text, in seconds. The Library of Congress, with 80 million items, contains an estimated 25 terabytes of information.

150 universities linked

Apple system puts 'reporter' on 'beat'

By Rory J. O'Connor
Mercury News Computing Editor

Apple Computer Inc.'s Rosebud project, still being developed in the company's research labs, is one illustration of how information services of the future might work.

The basic idea is familiar: a set of "reporters" — computer programs, actually — scour available information sources for data on their "beats." The data goes into a "notebook" from which the computer constructs a custom "newspaper" with a column devoted to abstracts of each reporter's findings.

For example, project engineer Kevin Tiene's computer has several reporters scanning his test data bases for information on Apple, Dow Jones and the Indianapolis 500 race. To create each one, he filled out a sort of form on the screen, first typing a question like "Who won the Indy 500?" and then checking off each data base the reporter should search. He also indicated if he wanted an automatic search, how frequently he wanted it and even how many "stories" the reporter should list.

Each reporter is represented on Tiene's screen by the icon of half a man's head in a fedora with a press card in its band. He can select them any time he wants to get the latest information on their beats.

Or he can just read the paper.

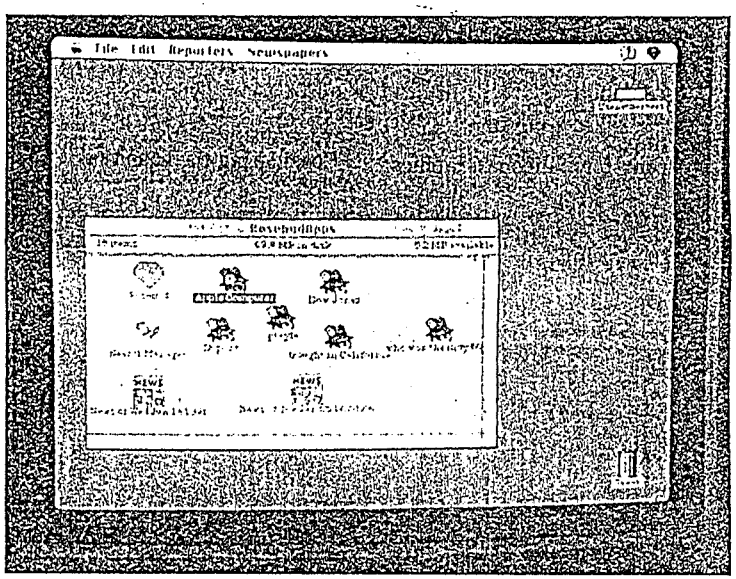
> The goal is to let users tailor their searches to their own needs.

Every day, each reporter automatically does its job and gives the computer abstracts of any text it finds. The computer then assembles the newspaper as a series of two-column pages on the screen, with one subject per column. Tiene can then read each abstract.

If he wants to see the full text of the "story," he simply points to it, and the computer locates the text on the network and shows it on the screen. If he wants to get a more detailed look at the reporter's notebook for past stories, he simply points to the reporter's icon above its column in the newspaper, and all the reporter's abstracts are displayed.

What Rosebud's reporters find won't necessarily be limited to text, Tiene says.

"The searching will all be done on a textual basis," he says, "but if you do a search on (Libyan leader Moammar) Gadhafi, you could get back a picture of him or even a video." That's because



data bases would include things like photo captions and the closed captions now included on many television programs.

But the 3-year-old Rosebud work hasn't solved some of the stickiest problems of such services, such as how information owners will be compensated for its use and even how ownership rights will be maintained. So far, the Wide Area Information Servers at Apple offer mostly information that's in the public domain, says Janet Vratny-Watts, a technology specialist in Apple's corporate library. "It doesn't have anything really compelling on it," she says.

"If this is going to be real, there are some serious issues of pricing and accounting to deal with," says Harvey G. Lehtman, who is in charge of the Rosebud project.

Tiene thinks data-base providers will have to abandon their current practice of charging fees for the time a user is connected to the data base, if for no other reason than computer reporters will be far more efficient searchers than humans. "They'll want to change to transaction-based charges," perhaps offering abstracts for little or no fee and charging significant fees when users ask for the whole document, he says.

The researchers also are concerned about security problems inherent in automated two-way computer communications, such as "crackers" using the library connection to infiltrate corporate computer systems.

150 universities linked

Already, an experimental computer library has linked 150 universities to 40 sources of information, ranging from National Institutes of Health data to corporate documents and Shakespeare's plays. New software allows users to browse or zero in on particular information.

As methods of retrieving information are standardized and perfected, industry executives and computer scientists say, thousands of new services, ranging from electronic newspapers to the computer equivalent of free public libraries, will blossom.

"Everyone is realizing how important it is to get into the mass market for information," said Thomas Koulopoulos, president of Delphi Consulting Group, a Boston market research firm.

Political disputes loom

Such ready access to huge amounts of computerized information has been the dream of many in the industry. But a lack of computing power, effective software and high-speed digital networks has stalled progress until recently.

If many of the technical problems are being solved, major business and political disputes remain. The researchers acknowledge that they must resolve several questions of privacy and pricing before they can put the new methods to commercial use.

Many sources of information, like government documents, might be available free, but other services, including electronic newspapers, will be available only to those who pay. The industry has yet to settle on ways to protect and charge for intellectual property in a computer net-

See LIBRARY, Page 5F

Cover Story

The world at your keyboard

Nationwide network links gobs of information

LIBRARY, from Page 1F

work where information can be copied instantly. But to encourage progress, Thinking Machines Corp., a Cambridge, Mass., supercomputer manufacturer, has made its software available free.

Some industry enthusiasts say the new technology will transform the way computerized information is sold. Mitchell Kapor, founder of Lotus Development Corp., predicts the growth of a new industry as significant as the personal computer business.

Some companies, like Dow Jones Corp. that provide computerized information over telephone lines have taken part in developing the new computer library.

In 1989, Thinking Machines enlisted the support of Dow Jones, Apple Computer Inc. and the KPMG Peat Marwick accounting and consulting firm to design the computer library, called Wide Area Information Servers, or WAIS. The system permits computer users to search through a huge volume of information quickly even if it is stored at several distant locations.

The system lets users conduct searches by typing common English phrases instead of more complicated computer commands.

While current systems like Dialog and Nexis require users to specify precisely the information they want, the new system can respond to a user's inferences. It initially presents a sample list of documents. The user chooses one or several, and then a "relevance feedback" program presents other documents most like the ones selected.

> Users search by typing common English phrases.

"This solves the problem of how to get to the information you need, getting not too much and not too little," said Esther Dyson, editor of Release 1.0, a computer industry newsletter.

This is a sharp contrast to the way services operate today, Dyson said. A computer user may need to call seven or eight data bases depending on the kind of information needed.

The WAIS system lets users of Apple computers harness a network of Thinking Machines supercomputers and smaller "server" computers to search data bases stored by Dow Jones, KPMG and several corporations and universities. Users also can read electronic mail, enter their corporate electronic libraries and summon up a wide variety of documents, newspapers and magazines.

At Thinking Machines, the WAIS system serves as a "corporate memory," allowing employees to retrieve memos, documents and other internal information. Employees who may not be working together can share expertise.

"If someone did something in Los Angeles and I'm sitting in San Francisco, I may not know about the work," said Robin Palmer, a senior manager at Peat Marwick.

WAIS delivers information over Internet, a collection of 2,600 high-speed public and private computer networks. This government-sponsored system of data highways is rapidly being improved and turned to commercial uses.

The market for software that allows the rapid retrieval of computerized text is small but growing, according to industry analysts.

In 1989, the United States had fewer than 60,000 users. By the next year, total sales were about $120 million. The Delphi Consulting Group expects the market to grow to 160,000 users and $235 million by 1992.

"Information-retrieval technology is starting to spread from supercomputers all the way down to personal computers," said Brewster Kahle, a Thinking Machines scientist who has led the WAIS experiment.

The WAIS system is built on a procedure for retrieving information developed by librarians who initially set out to computerize their card catalogs.

The procedure — known in the field as Z39.50 — now has the support of the Library of Congress, Apple, Sun Microsytems Inc., Next Inc., Dow Jones and Mead Data Central.

In the future, a special directory, or "white pages," will keep an up-to-date list of all the separate sources on the network.